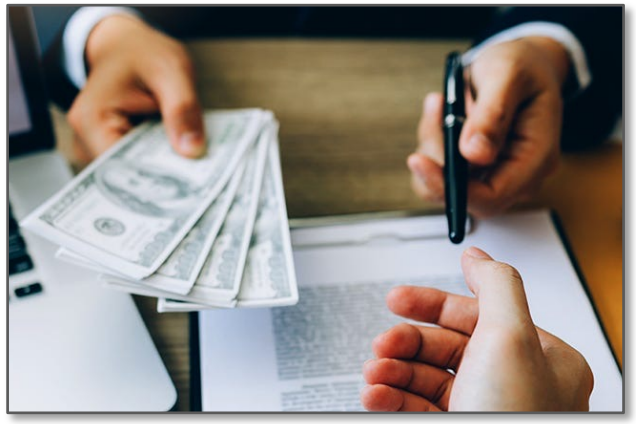
Strategic Representation

Vineet Nair, Ganesh Ghalme,
Inbal Talgam-Cohen, and **Nir Rosenfeld**
Technion CS

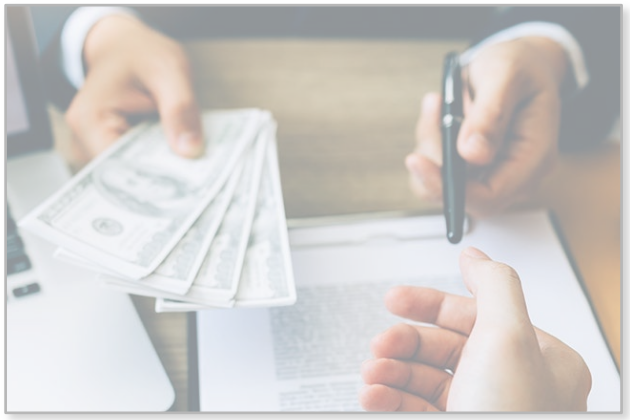
@ICML 2022

stratgic *classification*



users **game** learning system

stratgic *classification*

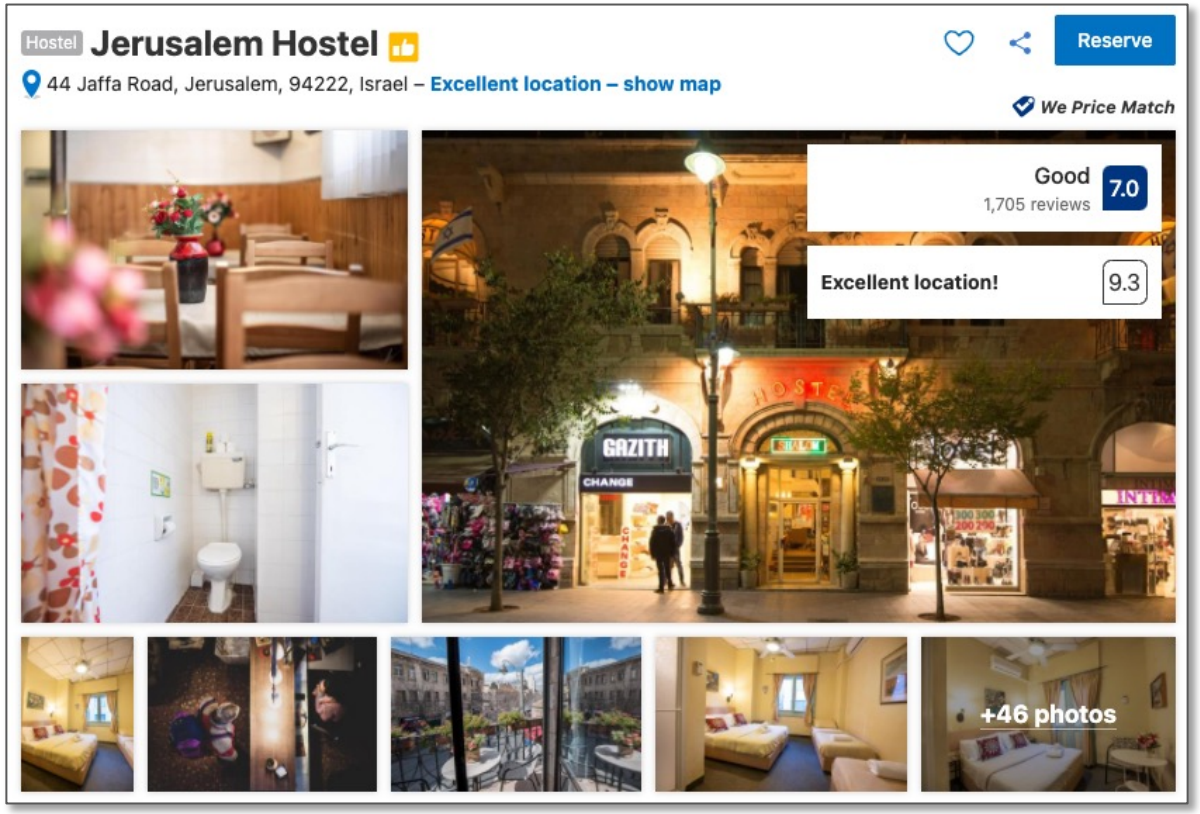


users **game** learning system

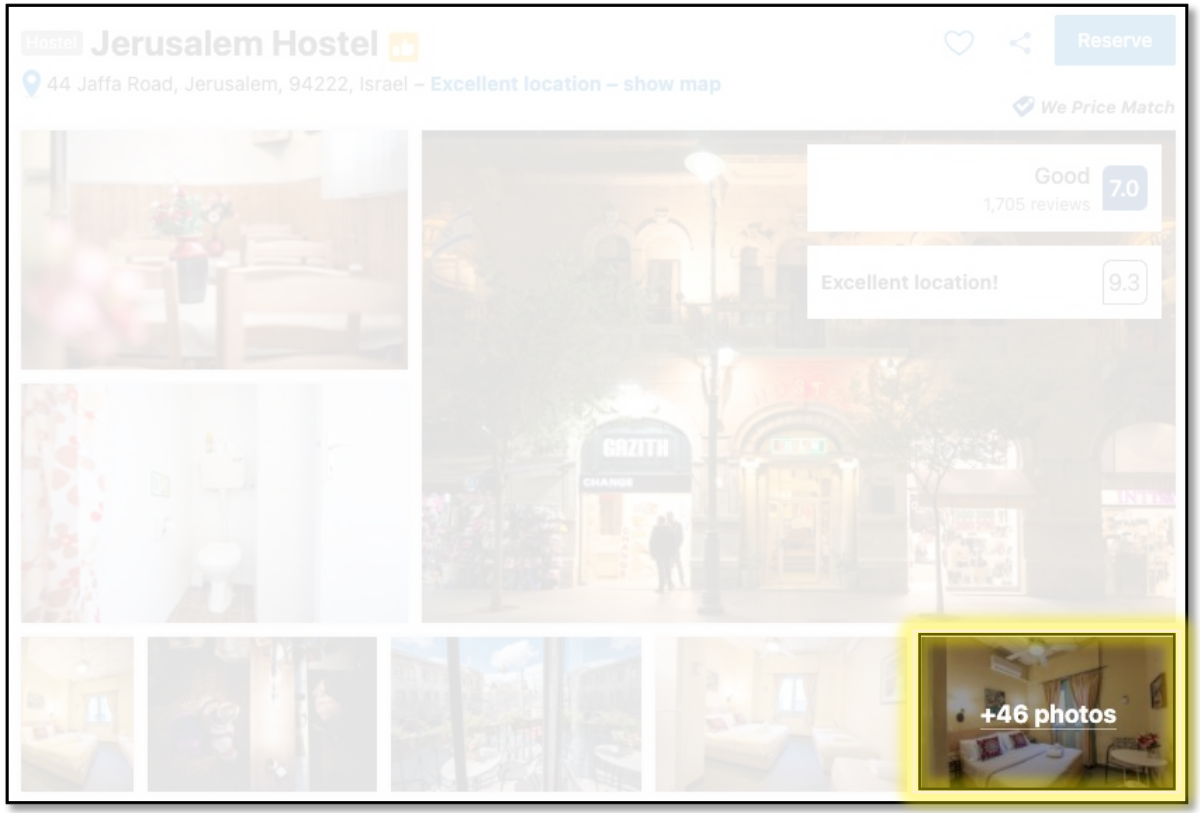
stratgic *representation*

system **games** learning users

representation:

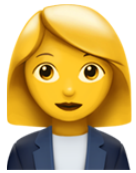


representation:



who chooses?
and why?

ideal:



choice function: $y = h^*(x)$ ← *choose iff worthwhile*

item: $x = \{$ 🛏️, 🛁, 🚪, 🍜, 🍷, 🌅, ⛩️, 🏖️, 🥇 $\} \in 2^E$

value: $y \in \{+1, -1\}$

reality:

choice function: $\hat{y} = h(z)$ ← *choose iff worthwhile*

representation: $z = \{ \text{🛏}, \text{🛁}, \text{🪜}, \text{🍜}, \text{🍷}, \text{🟥}, \text{⛩}, \text{⛱}, \text{🥇} \} \subseteq x, \ |z| \in [k_1, k_2]$

truthful, but lossy

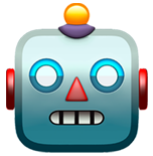
value: $y \in \{+1, -1\}$

strategic

reality:

choice function: $\hat{y} = h(\phi_h(x))$ ← *choose iff worthwhile*

strategic representation: $\phi_h(x) = \underset{z \subseteq x, |z| \in [k_1, k_2]}{\operatorname{argmax}} h(z)$



value: $y \in \{+1, -1\}$

just choose!

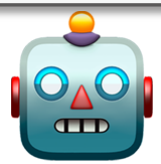
strategic

reality:

learned

choice function: $\operatorname{argmax}_{h \in H} \mathbb{E}_{(x,y)\sim D}[\mathbb{1}\{h(\phi_h(x)) = y\}]$ ← *choose iff worthwhile*

strategic representation: $\phi_h(x) = \operatorname*{argmax}_{z \subseteq x, |z| \in [k_1,k_2]} h(z)$



just choose!

value: $y \in \{+1, -1\}$

strategic

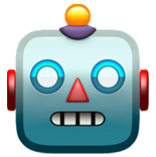
reality:

learned

choice function: $\text{argmax}_{h \in H} \mathbb{E}_{(x,y)\sim D}[\mathbb{1}\{h(\phi_h(x)) = y\}]$

$\hat{h}$ = *"foodie"*

strategic representation: $\phi_{\hat{h}}(x) = \{$ 🛏️, 🛁, 🍜, 🍷, ⛩️, 🥇 $\}$



most enticing truthful representation

strategic

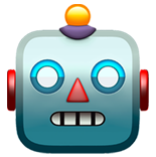
reality:

learned

choice function: $\text{argmax}_{h \in H} \mathbb{E}_{(x,y)\sim D}[\mathbb{1}\{h(\phi_h(x)) = y\}]$

$\hat{h}$ = *"tourist"*

strategic representation: $\phi_{\hat{h}}(x) = \{$ 🛏️, 🛁, 🈵, 🍜, 🍷, 🌅, ⛩️, 🏖️, 🥇 $\}$



most enticing truthful representation

strategic

reality:

learned

choice function:

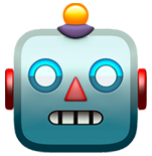
$$\operatorname{argmax}_{h \in H} \mathbb{E}_{(x,y)\sim D}[\mathbb{1}\{h(\phi_h(x)) = y\}]$$

choose iff worthwhile ←

strategic representation:

$$\phi_h(x) = \operatorname*{argmax}_{z \subseteq x, |z| \in [k_1, k_2]} h(z)$$

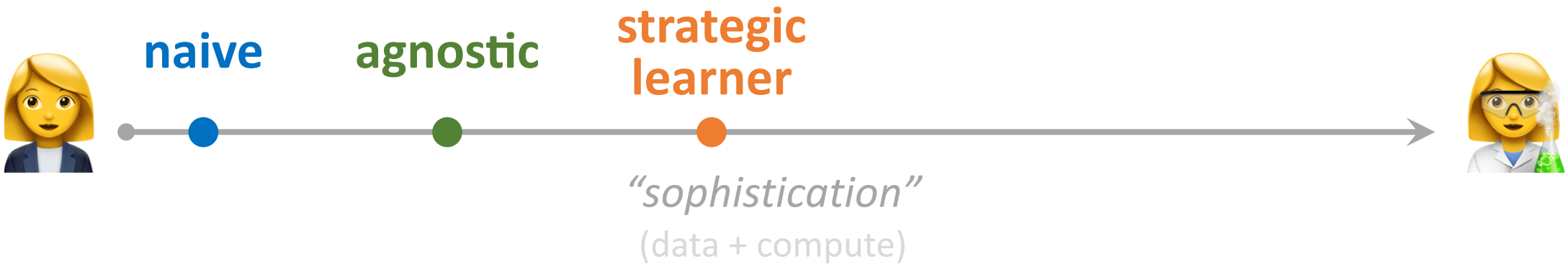
just choose!



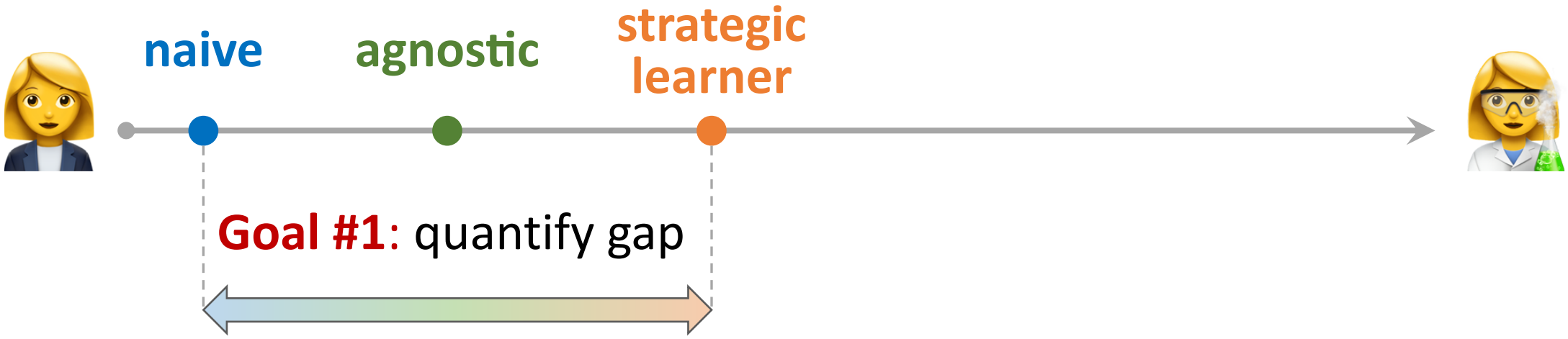
goal: learn accurate, strategy-robust choice function $\hat{h}$ ← *possibly* $\neq h^*$

focus:
study how users cope with
strategic representations

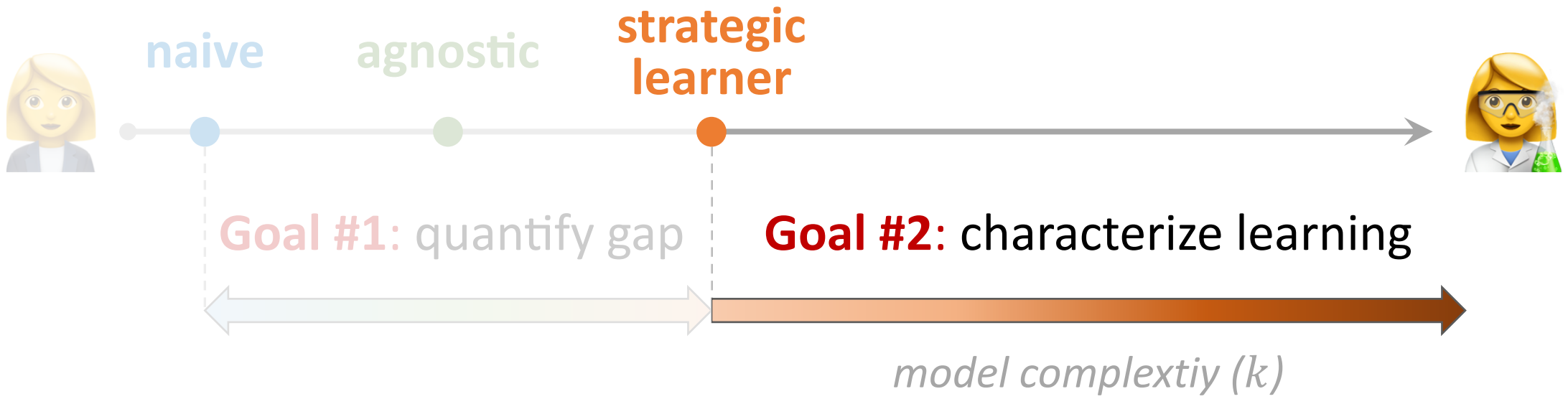
focus:
study how users cope with
strategic representations



focus:
study how users cope with
strategic representations



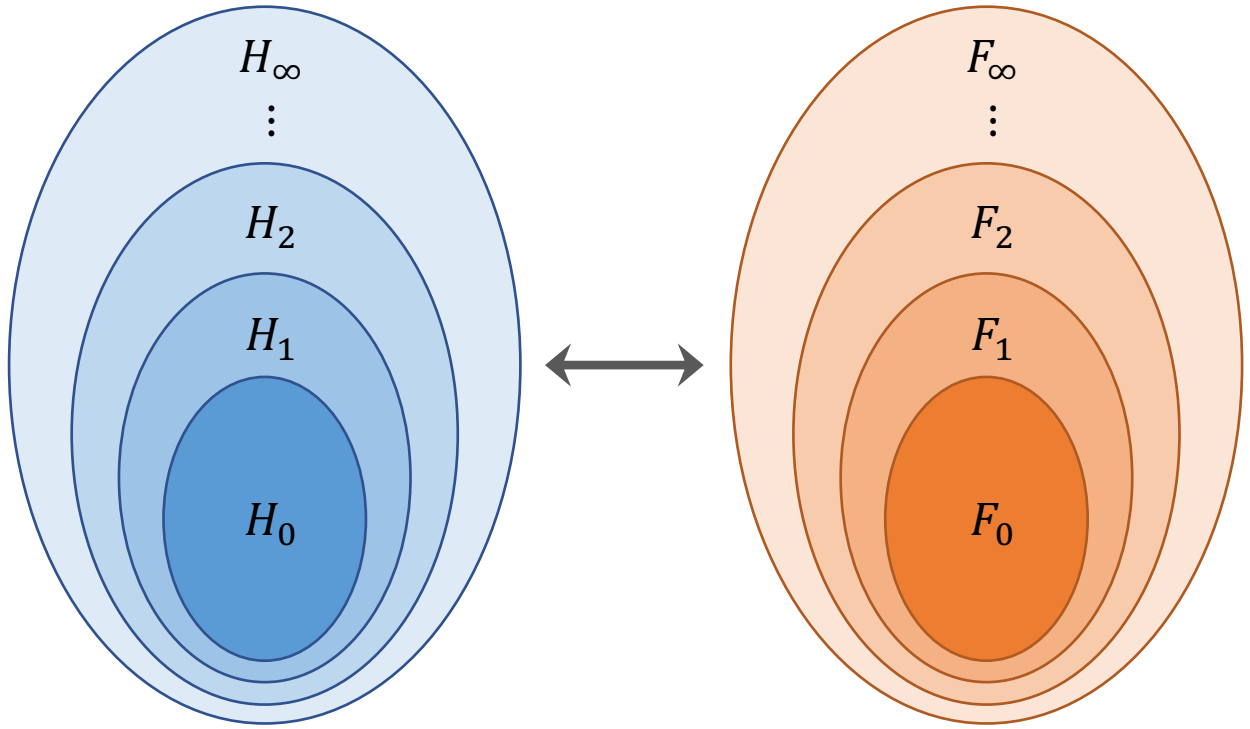
focus:
study how users cope with
strategic representations



Results and contributions:

1. Structural foundations

- Choice function class hierarchy: $H_k \subset H_{k+1}$
- Induced functions: $f(x) = h(\phi_h(x))$
- Shared structure



Results and contributions:

1. Structural foundations

- Choice function class hierarchy: $H_k \subset H_{k+1}$
- Induced functions: $f(x) = h(\phi_h(x))$
- Shared structure

2. Strategy-robust learning algorithm

efficient and exact

(under "realizability")

Results and contributions:

1. Structural foundations

- Choice function class hierarchy: $H_k \subset H_{k+1}$
- Induced functions: $f(x) = h(\phi_h(x))$
- Shared structure

2. Strategy-robust learning algorithm

efficient and exact

(under "realizability")

3. "Balance of Power" analysis

- **User's perspective**: expressivity vs. effort (k)
  approximation error + estimation error (gen. bound)
- **System's perspective**: exogeneous constraints (k_1, k_2)

Thanks!

(come to our poster)

check out our other paper on:
generalized strategic classification
and the **case of aligned incentives**
(also @ICML2022)